LA-UR- 00-5972

Title: Portland Synthetic Population

Author(s): Brian W. Bush

Submitted to: Transportation Research Board 80th Annual Meeting
7-11 January 2001, Washington, D.C.

Portland Synthetic Population

B. W. Bush and the TRANSIMS team
Los Alamos National Laboratory
8 January 2001

Abstract

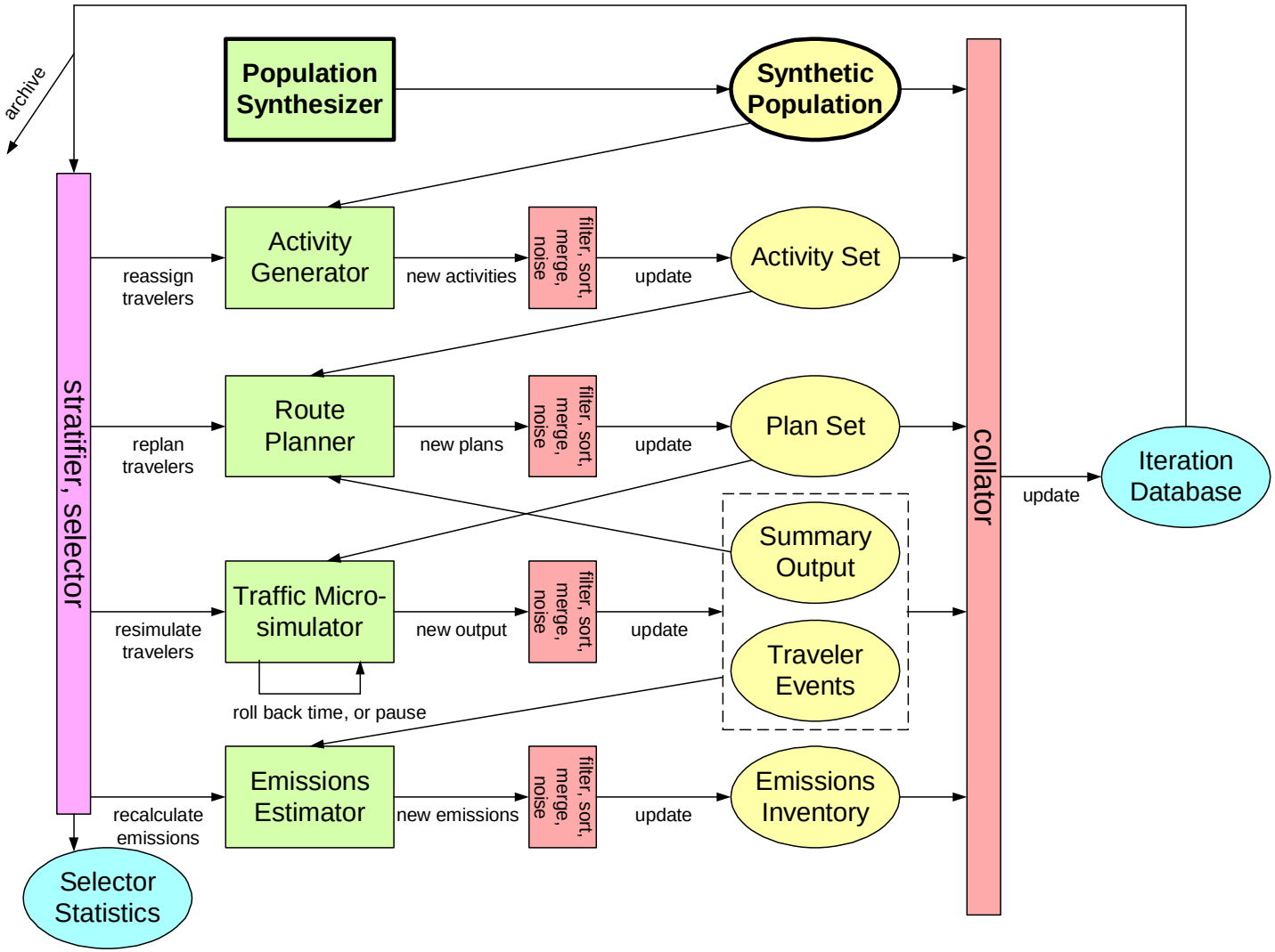
TRANSIMS (Transportation Analysis and Simulation System) is an integrated system of travel forecasting models designed to give transportation planners accurate, complete information on traffic impacts, congestion, and pollution. The Population Synthesizer Module constructs a regional population imitation with demographics closely matching the real population. This presentation outlines how we have constructed the synthetic population for our Portland, Oregon, case study.

Los Alamos National Laboratory is leading this effort to develop these new transportation and air quality forecasting procedures required by the Clean Air Act, the Intermodal Surface Transportation Efficiency Act, and other regulations; it is part of the Travel Model Improvement Program sponsored by the U.S. Department of Transportation, the Environmental Protection Agency, and the Department of Energy.

Outline

- *approach*
  - *purpose*
  - *data flow*
  - *algorithm*
- *application in Portland*
  - *network*
  - *population forecast*
  - *correlation structure*
  - *quality of fit*
  - *demographics*
  - *household location*
  - *vehicle assignment*
- *conclusion*

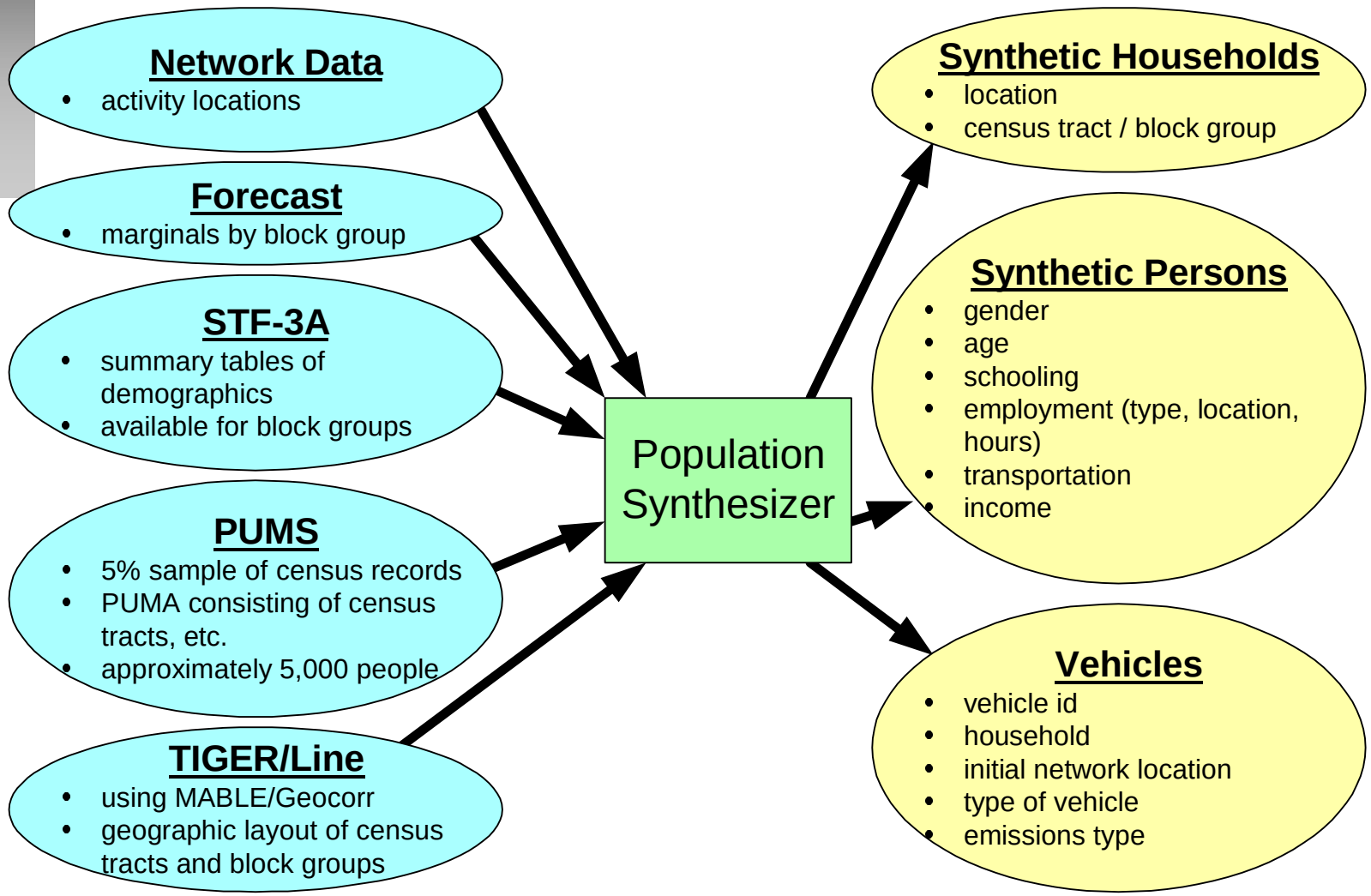
TRANSIMS Architecture



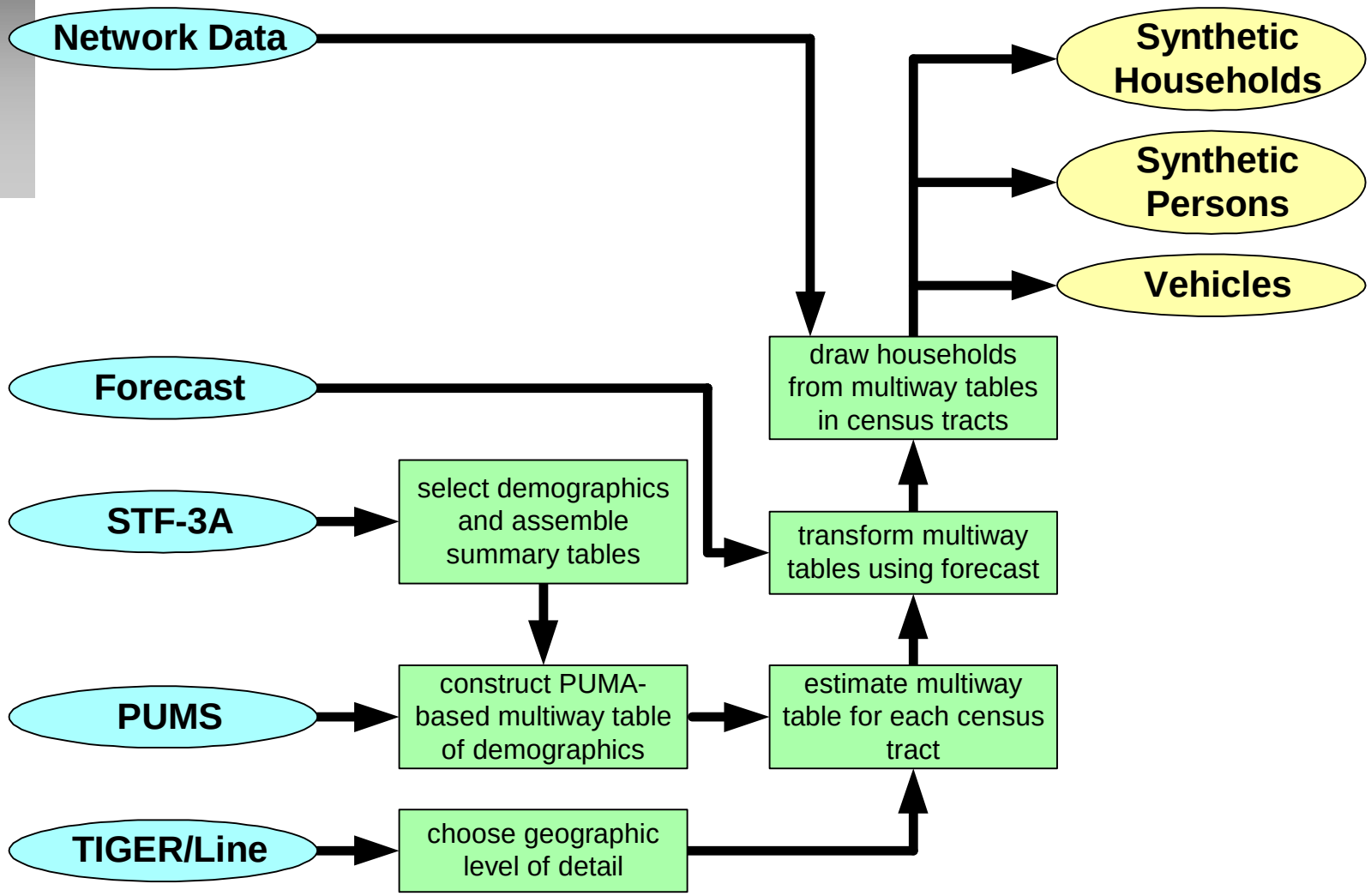
Purpose

- *creates a regional population imitation*
  - *demographics closely match real population*
  - *households are distributed spatially to approximate regional population distribution*
  - *household locations determine some of the travel origins and destinations*
- *synthetic population's demographics form basis for individual and household activities requiring travel*

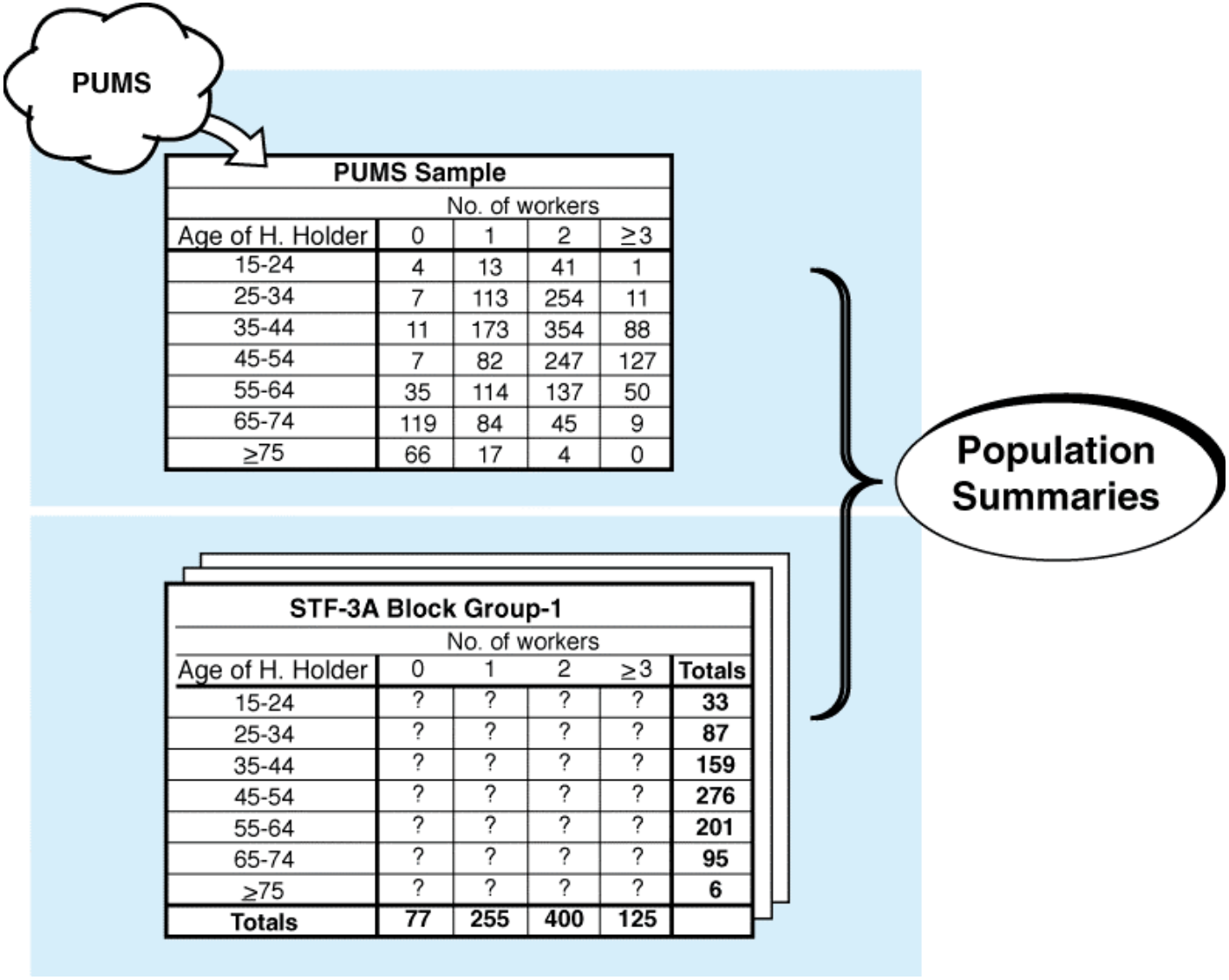
Data Flow


Network Data
activity locations
Forecast
marginals by block group
STF-3A
summary tables of demographics
available for block groups
PUMS
5% sample of census records
PUMA consisting of census tracts, etc.
approximately 5,000 people
TIGER/Line
using MABLE/Geocorr
geographic layout of census tracts and block groups
Population Synthesizer
Synthetic Households
location
census tract / block group
Synthetic Persons
gender
age
schooling
employment (type, location, hours)
transportation
income
Vehicles
vehicle id
household
initial network location
type of vehicle
emissions type

Algorithm


Network Data
Synthetic Households
Synthetic Persons
Vehicles
draw households from multiway tables in census tracts
Forecast
STF-3A
select demographics and assemble summary tables
transform multiway tables using forecast
PUMS
construct PUMA-based multiway table of demographics
estimate multiway table for each census tract
TIGER/Line
choose geographic level of detail

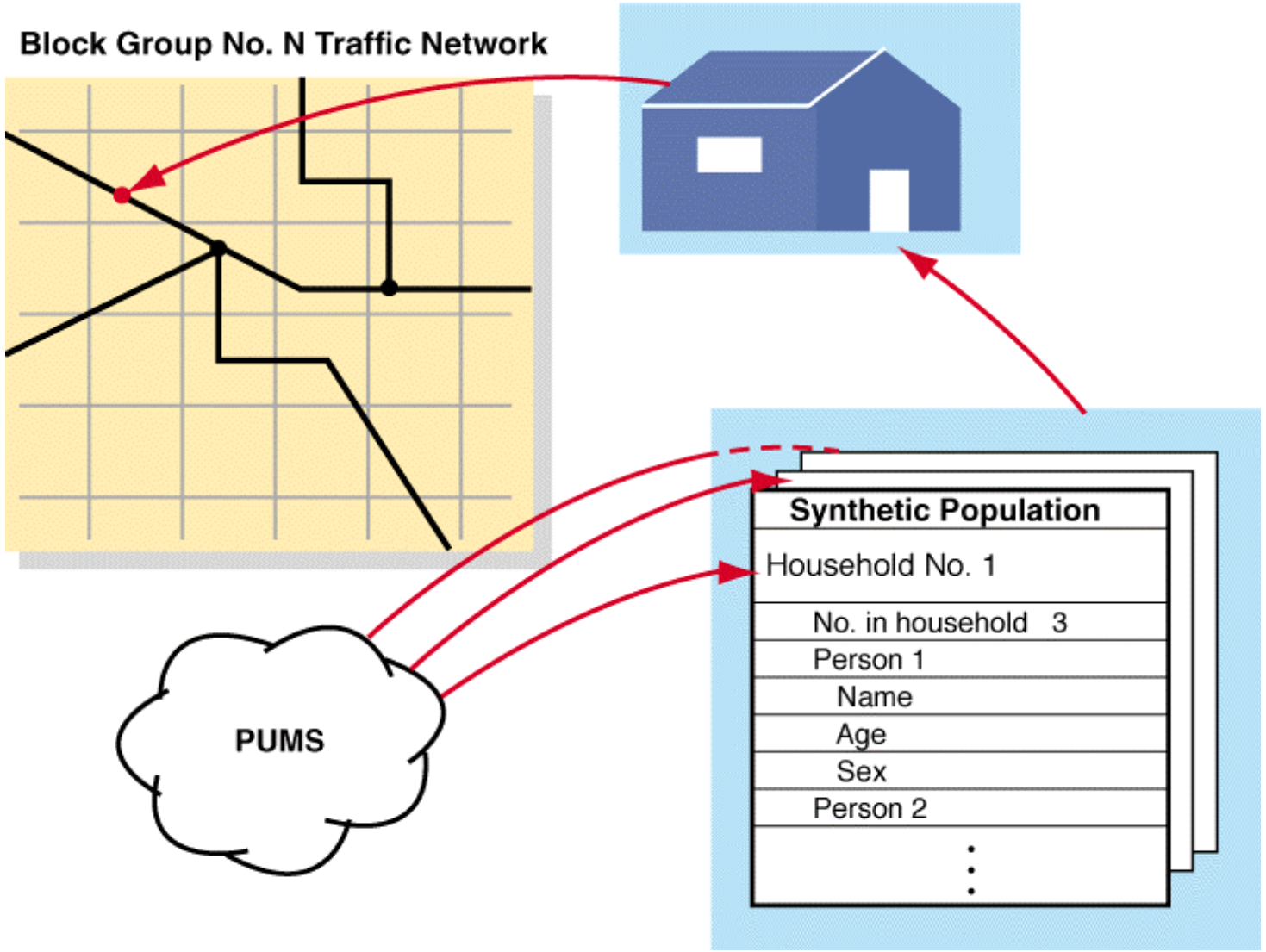
Disaggregating Census Data



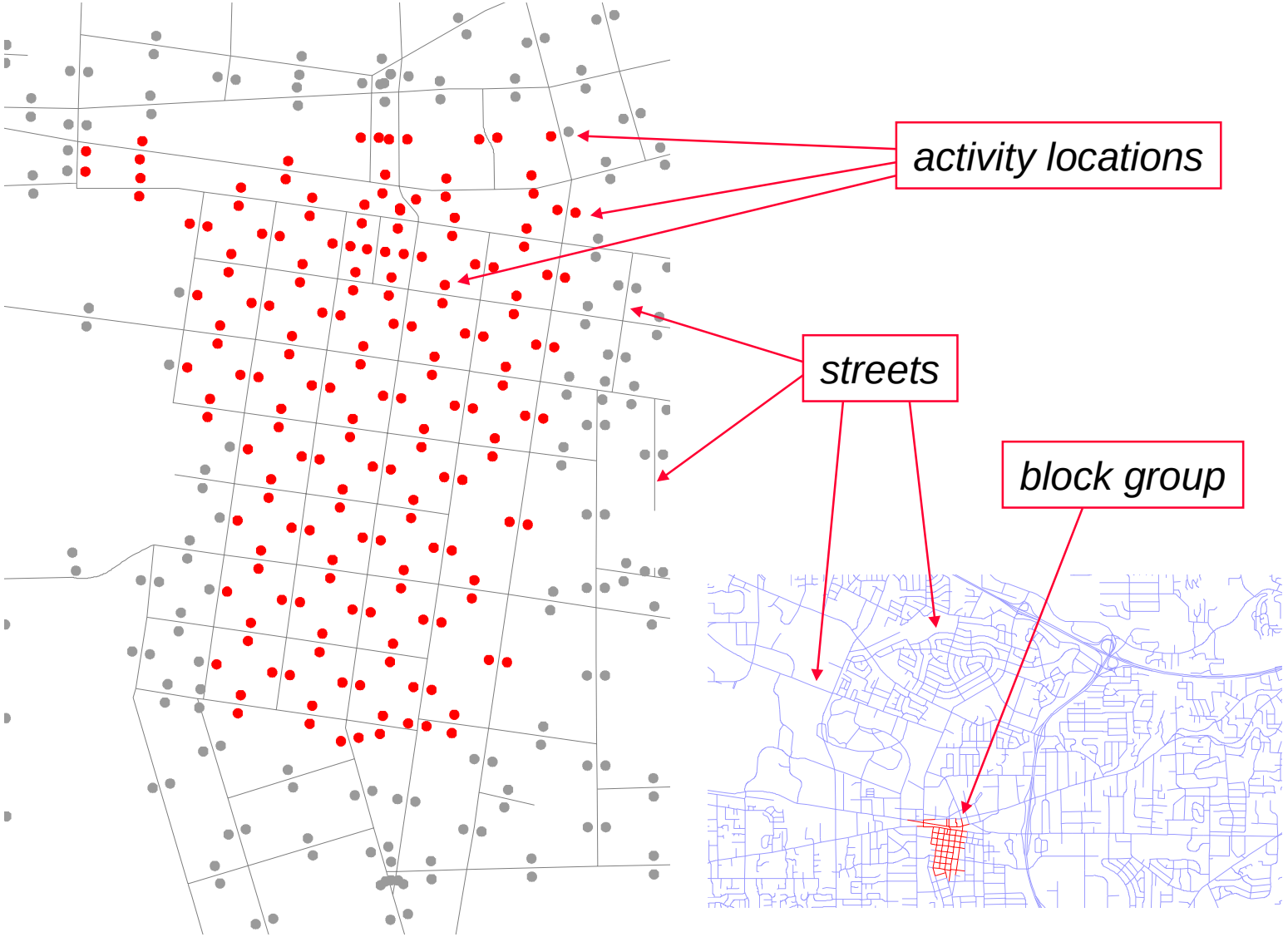
Choice of Variables from STF-3A Summary Tables

- *family households*
  - *age of the householder [P24]*
  - *family income [P107]*
  - *number of workers in the family [P112]*
  - *poverty status × race × family type × presence and age of children [P124A,B]*
- *non-family households*
  - *household type and gender [P17]*
  - *race × household type × presence and age of children [P20]*
  - *age of non-family householder [P24]*
  - *non-family household income [P110]*
  - *poverty status × age of householder × household type [P127]*
- *group quarters*
  - *group quarters [P40]*
  - *group quarters × age [P41]*

Located Population



Block Group 31200.1 in Portland, Oregon



1996 Forecast for Block Group 31200.1

- *any forecasting methodology may be used*
- *forecast represented as a marginal distribution over block groups of . . .*
  - *household size*
  - *age of head of household*
  - *annual household income*

Size	*Households*
1	84
2	42
3	0
4	6
5	0
6	0
⩾ 7	0
Total	*132*

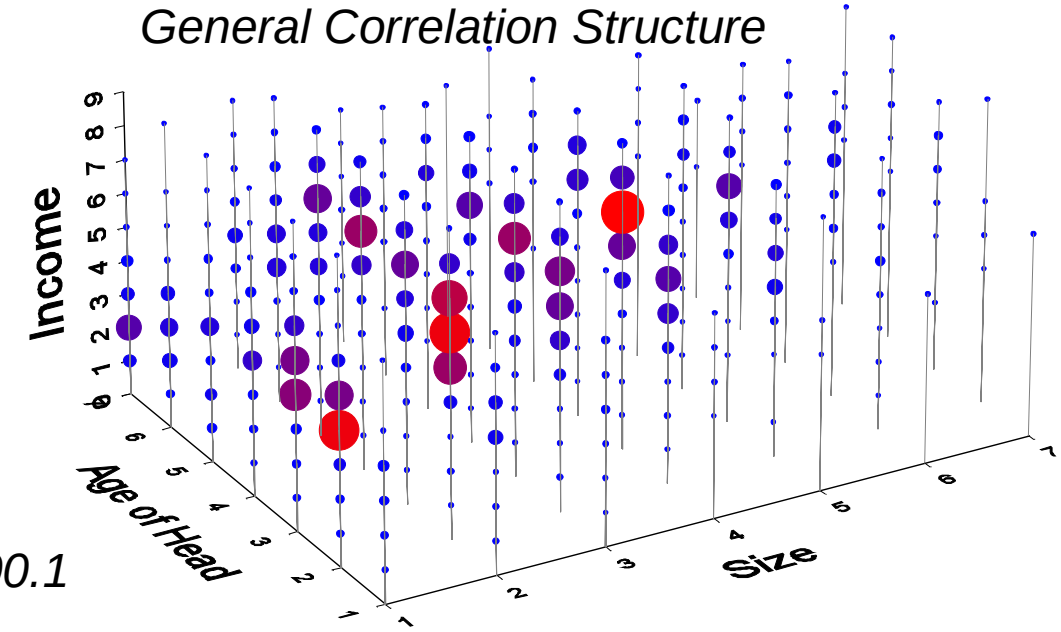
Age of Head	*Households*
⩽ 24	24
25–34	42
35–44	35
45–54	28
55–64	0
65–74	2
⩾ 75	1
Total	*132*

Income	*Households*
⩽ 4999	0
5000–9999	47
10,000–14,999	8
15,000–24,999	19
25,000–34,999	37
35,000–49,999	0
50,000–74,999	21
75,000–99,999	0
⩾ 100,000	0
	132

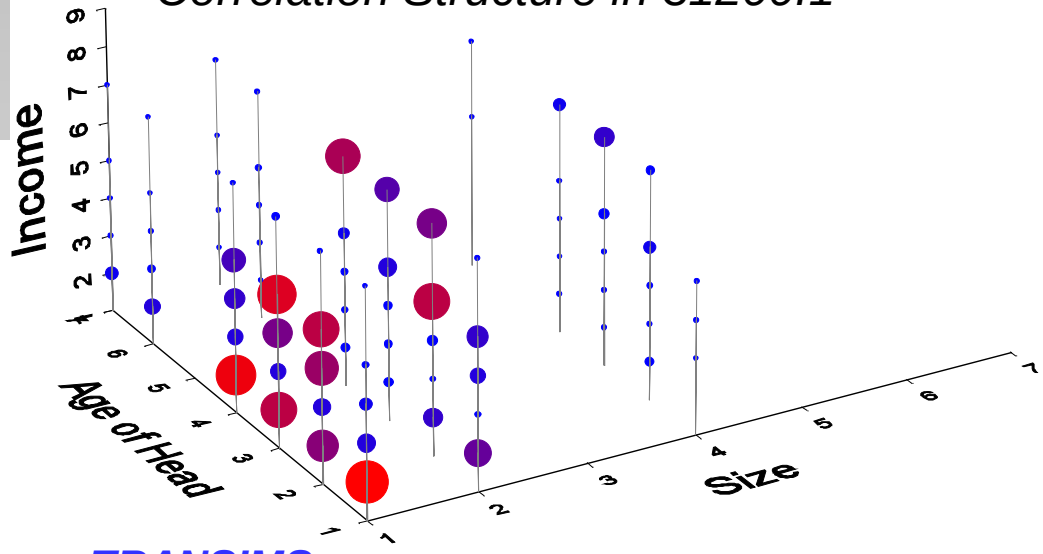
Iterative Proportional Fitting for Block Group 31200.1

- *correlation structure of demographic variables preserved*
- *marginal distributions of forecast matched*

General Correlation Structure



Correlation Structure in 31200.1

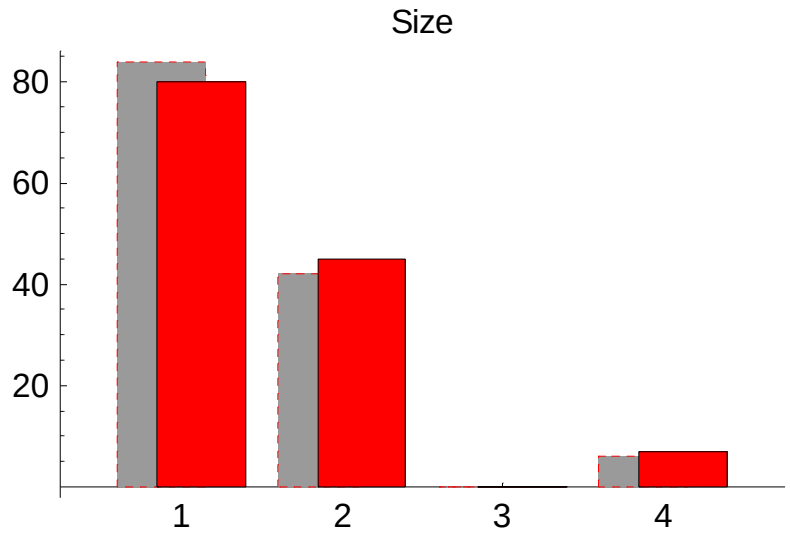


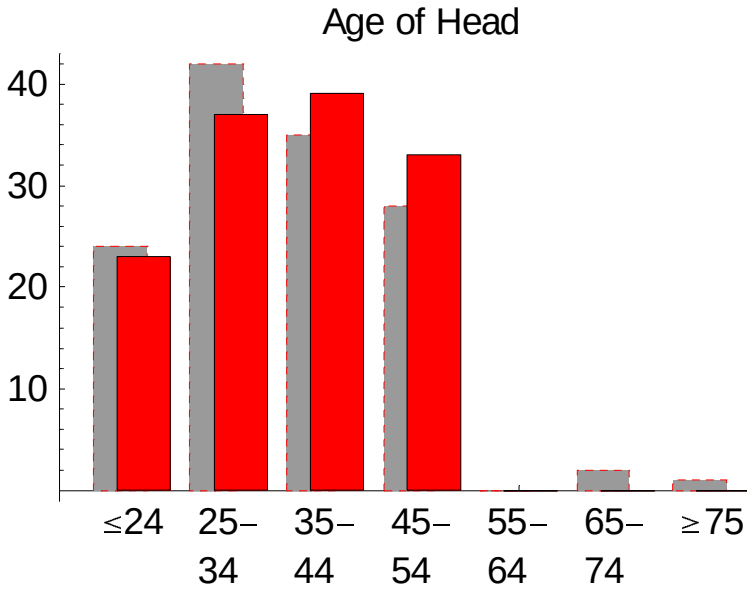
- *correlation structure measured by "odds ratio," e.g.,*

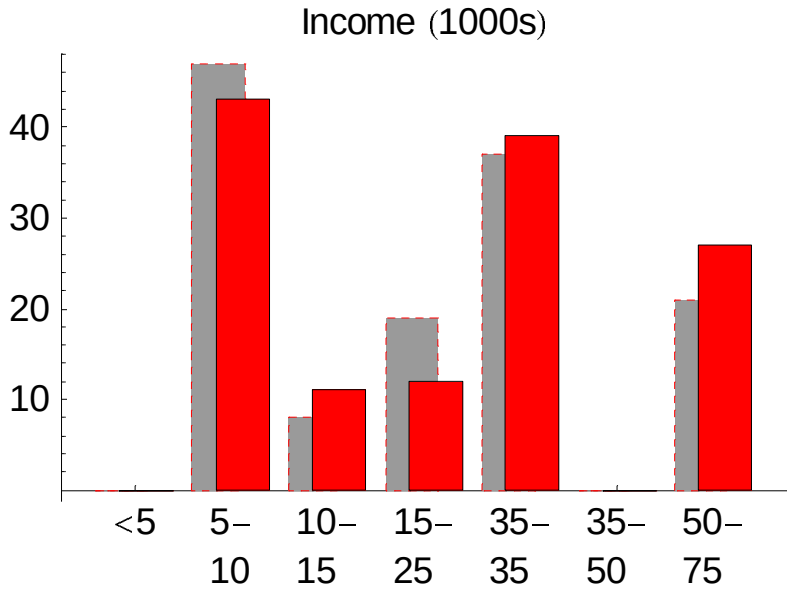
$$\varphi = \frac{p_{1,1} \cdot p_{2,2}}{p_{1,2} \cdot p_{2,1}}$$

Forecast vs. Synthetic Population in Block Group 31200.1

- *marginal distribution of demographic variables closely matched:*
  - *synthetic population (red)*
  - *forecast (gray)*







Example Household in Block Group 31200.1

Household

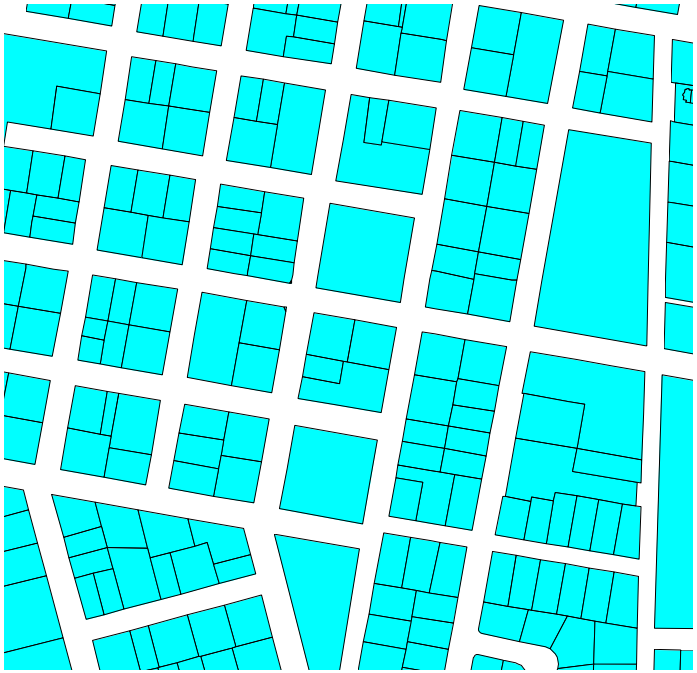
ID	111733
Size	4
Vehicles	3
Activity Location	23101
PUMS Record	44789
Anyone under 18	Yes
Workers in 1989	3+
Total Income	$64,000
Tenure	Owned with mortgage or loan
Value	$90,000 - $99,999

Persons

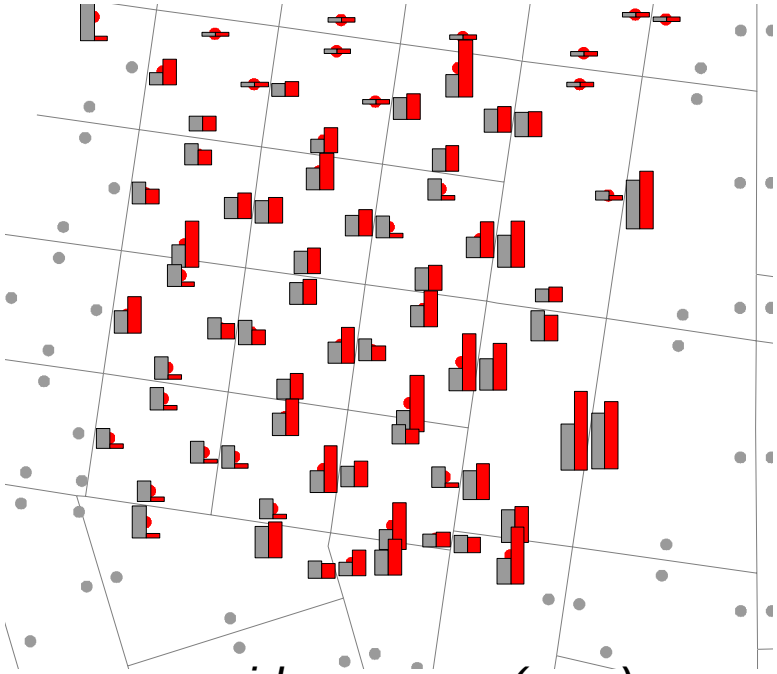
ID	255552	255553	255554	255555
Age	42	42	19	7
Relationship	Householder	Husband/wife	Son/daughter	Son/daughter
Sex	Male	Female	Female	Female
Worked in 1989	Yes	Yes	Yes	No (under 18)
Educational Attainment	Some college, but no degree	High school graduate, diploma or GED	Some college, but no degree	1st , 2nd, 3rd, or 4th grade
Industry	Electrical Machinery, Equipment, and Supplies, N.E.C	Not Specified Retail Trade	Offices and Clinics of Chiropractors	
Occupation	Managers and Administrators, N.E.C	Sales Workers, Other Commodities	Managers, Medicine and Health	
Total Income	$45,000	$13,000	$6000	
Hours Worked	40	40	15	
Lived Here in 1985	No	No	No	(under 5)
Means of transportation to work	Car, truck, or van	Car, truck, or van	Car, truck, or van	
Vehicle occupancy	1	1	1	
Time of departure for work	6:50	1:00	14:00	
Travel time to work	0:20	0:15	0:10	

Location of Households in Block Group 31200.1

- *density of residences along streets provided by Portland tax lot data*
- *households placed at activity locations proportionally to area of residences along street*
- *households not placed in commercial, industrial, etc. areas*



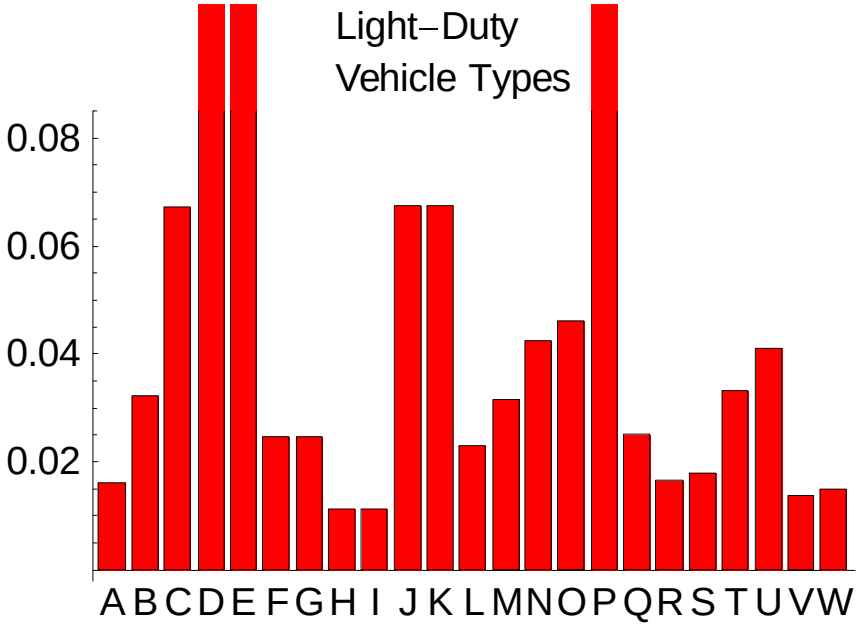
tax lots

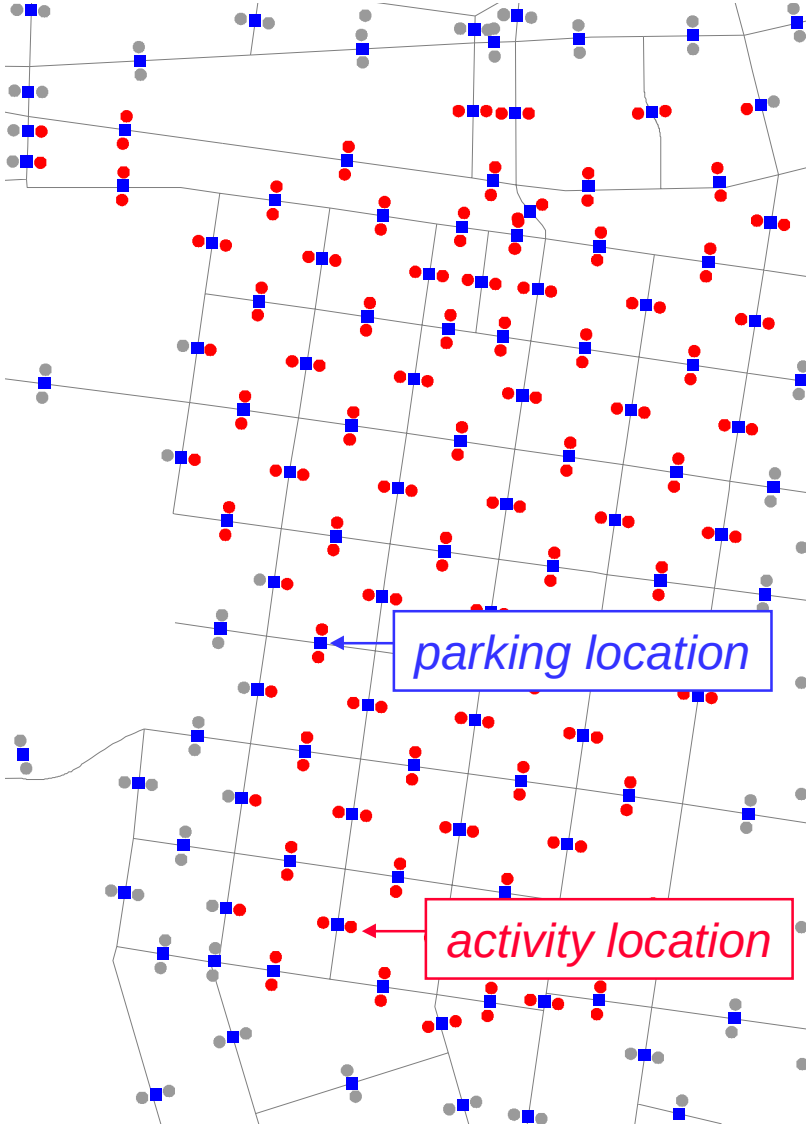


residence area (gray)
synthetic household count (red)

Vehicle Assignment

- *vehicles assigned according to household demographics*
- *vehicles parked at location nearest household*
- *distribution of 23 CMEM vehicle types determined from Oregon motor vehicle records*





Conclusion

- *statistically realistic synthetic population*
  - *correlation structure preserved*
    - *joint distributions of demographic variables*
    - *spatial distribution of households*
    - *number of vehicles for households*
  - *forecast marginal distribution matched*
- *detailed demographic information available on individual basis*
  - *65 household variables*
  - *74 person variables*
- *practical input data requirements*
  - *census (STF-3A, PUMS, TIGER)*
  - *land use (e.g. tax lots or zoning)*
  - *vehicle (e.g., motor vehicle records)*
  - *population forecast*